DOCUMENT RESUME

ED 172 804 IR 007 439

AUTHOR Winn, Bill
TITLE Field-Testing Instructional Materials.
PUB DATE Mar 79
NOTE 23p.; Paper presented at the Annual Convention of the Association for Educational Communications and Technology (New Orleans, Louisiana, March 1979); For related document, see IR 007 416

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Affective Behavior; Attitudes; Audiovisual Aids; *Evaluation Methods; *Formative Evaluation; *Instructional Materials; Student Developed Materials; *Task Analysis; *Task Performance

ABSTRACT

This study examines simple techniques for field testing instructional materials, and is intended for the educational practitioner who is not trained in instructional development procedures and does not have the time or money for elaborate evaluation procedures. Described are critera for field testing instructional effectiveness, procedures for specifying task descriptions relating to instructional programs, and testing procedures for the tasks involved. A sample instructional task--laminating a picture--is included, illustrating the steps in defining the task, evaluating for task completion, and testing the affective component of the instructional method. A list of references and two figures are attached. (RAO)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED172804

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

FIELD-TESTING INSTRUCTIONAL MATERIALS

by

Bill Winn
University of Calgary

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Michael Simonson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Paper presented at the annual convention of the Association for Educational Communications and Technology, New Orleans, March 1979.

IR007439

Abstract

The application of research and theory to the development of instructional materials relies largely for its success upon testing materials in the field and revising them accordingly. This need not necessarily be a complicated business. The practitioner who is not a trained instructional developer, can test the effectiveness, feasibility and the affective component of instruction. This paper describes a few simple methods for doing so.

In 1972, AECT's Committee on Definition and Terminology published a definition of educational technology, (1972). The basis of this definition was a statement by John Kenneth Galbraith, from *The New Industrial State,* that "Technology is the systematic application of scientific or other organized knowledge to practical tasks", (1967, p.12). The three components of this statement, systematic application, organized knowledge, and practical tasks, describe the relationship that exists between research on the one hand, and instructional development on the other, between the theory and the practice of educational technology. AECT's definition can be interpreted to mean that educational technology is a three-phase problem-solving process whereby educational problems are identified, appropriate knowledge is marshalled that is likely to provide a solution to these problems, and whereby this knowledge is systematically applied so that the solution is both useable and effective. Instructional development is pivotal in this process.

If we look a little more closely at each of the three components of Galbraith's statement, the nature of the development process, and the relationships between the various roles of the educational technologist become clearer. We must first ask what are the practical tasks of educational technologists. Within the limited context of the activities of researchers and instructional developers, the answer to this question must be: to bring about situations in which learning can effectively occur. In practice, this means that the instructional developer is told of instructional need that has to be met. A need is usually taken to be "a measurable discrepancy between what is and what should be" (Kaufman, 1972); so the role of the instructional developer becomes one of bridging a gap,

with theory, research and experience as his bridging materials. To take a simple example, if a school board discovers that its high-school graduates cannot read, then a need has been identified: they are illiterate (what is); they should be able to read novels, magazines, and newspapers without difficulty (what should be). This gap is measurable insofar as such factors as working vocabulary size, reading speed and reading habits can be assessed empirically. The instructional developer's problem, then, is to develop instructional materials, strategies, environments which make the need disappear.

One place the instructional developer can turn for a solution to problems is to the body of research and theory that has been accumulated over the years. Here may be found information about how developers solved similar problems elsewhere, which can be applied directly, or more general psychological theories of how people read, why students can get through school and still be illiterate, and the best ways of teaching reading. The developer must also look at the type of learner in question. Are they motivated? Are they of high, average or low ability? The feasibility of the various strategies that begin to emerge as possible ways of solving the problem must also be taken into account. The developer, in effect, is gathering a large amount of information about possible solutions to the problem and about the particular context in which the problem has presented itself. This information gathering, through learner analysis, task analysis, feasibility analysis, and the provision of several potential solutions, is a critical step. Solutions to instructional problems can never be better than the information on which they were based, and they are frequently worse.

It is in the systematic application of this information to a

particular problem that the instructional developer must excel. The application of theory to practical problems is systematic if it meets two criteria: first, the theory, experience and research that are applied must be pertinent; second, the decisions arising from the application of particular research and theory must be empirically tested. The second point is extremely important. It does not matter how "correct" or relevant research and theory appear to be. They will never be perfectly appropriate to the particular circumstances in which the instructional developer has to apply an instructional solution. After all, tests of statistical significance are never made at the 100 percent level. And the generalizability, that characterizes theory, implies the need for adaptation to specific situations. So however sure the instructional developer is that the right instructional decisions have been made, the instructional products must nevertheless be tried out in the field and revised until predetermined standards are achieved at which point the developer will be sure that the need has been met. The rest of this paper addresses itself to this question.

Evaluation

Before getting into the "how to" of field testing, it is necessary to dwell for a moment on the more general question of evaluation. It is customary to distinguish between two general types of evaluation: formative and summative (Bloom, Hastings and Madaus, 1971). The difference between the two is described by Gagné and Briggs (1974) in terms of what each type is used for. In their words, formative evaluation is used to "form" instruction; summative evaluation is used to determine the "summed" effects of instruction on learners. The main purpose of formative evaluation is

to "debug" early versions of instructional materials and activities, to see if they work, to identify weaknesses, to find out why they do not work, with a view to using the information gathered by evaluation to revise the materials and activities in ways that correct those problems. Information produced by formative evaluation is prescriptive. The purpose of summative evaluation is to gather information that is descriptive. Summative evaluation is carried out on larger units of instruction that are in some sense complete. The information gathered tells us how the finished package performs in normal instructional settings, and is often used to validate instruction (Thiagarajan, 1971). This paper is concerned with formative evaluation.

So that it is useful to the instructional developer (this term applies to anyone who is developing instructional materials, not necessarily a professional developer), formative evaluation of instruction must provide information on the effectiveness, feasibility and affective qualities of the instruction under development. While there is a tendency for field testing to get complicated, each of these three kinds of information can be gathered easily. Yet even simple methods must allow the developer to do more than simply conclude that the instruction does not work, cannot be implemented, or that the teachers and students do not like it. The assessment of the effectiveness of instruction under development must not only evaluate how much students learn, but must also permit the developer to pinpoint what the student does not learn well, and why. The purpose of formative evaluation, remember, is to guide revision of materials. It is therefore important to know not just that revisions are needed, but where they are needed. Likewise, formative evaluation must identify any reasons for infeasibility of implementation. If teachers say that they cannot

manage to carry out their part in the proposed instructional process, the developer needs to know why so that feasible alternatives can be tried next time around. Similarly, the affective component of formative evaluation must determine why students like some parts and not others, which parts are boring and which are exciting, which are interesting and which dull. Only then can the portions learners like less be made more palatable.

Evaluation in each of these three areas is only practical if a level of acceptable performance for the instruction is determined beforehand. After all, even the best designed instruction can never be one hundred percent perfect. It would therefore be a waste of the developer's time to keep on revising materials until all students scored full marks, or all teachers could take part, or everyone liked them. It is safe to say that the earliest revisions are the easiest to make since they deal with the most obvious errors and consequently will bring about the most marked improvements. Each subsequent round of field testing and revision will require increasingly finer adjustments to be made until a point is reached where the improvement in instruction brought about by the revision is so small that it is not to be worth the effort (and cost) to make it. While the developer can usually tell when this point has been reached, it is of considerable benefit to try to anticipate when the point of balance between what is effective and what is worthwhile will be arrived at. This is done by setting up criteria for acceptability. For example, it might be decided that no further revisions will be carried out when all learners score eighty percent or more on a posttest, or when a simple majority of teachers judge the new instruction to be feasible, or when no more than ten percent of the learners say that they do not like what has been developed. Sometimes these criteria are set by the developer's client, as in the case

of performance contracting. But this need not concern us now. Suffice it to say that criteria, set beforehand, often help the developer make the decision that further revisions are not necessary.

The rest of this paper describes a few simple yet perfectly adequate techniques for field-testing instructional materials. They are not intended for fully-fledged instructional developers, who have their own repertoire of far more sophisticated tests and methods. They are intended, rather, for the eduational practitioner, who is not a trained or experienced developer, who has neither the time nor the money to carry out elaborate evaluation studies, and whose development activities are modest in scope. The methods to be described have been gathered from far and wide, and for the most part used successfully by staff in the Learning Technology Unit at the University of Calgary (Learning Technology Unit, 1979).

Field Testing Instructional Effectiveness

Criteria. Typical criteria for the acceptance of instruction are:

i. Percentage performance on a posttest, ranging from a low of 70% to a high of 90% depending on the task or content.

ii. Number of students making errors, which is particularly useful for debugging instruction in motor skills.

iii. Time taken for students to complete a task on which 100% success is expected, such, again, as performing simple motor skills, or straightforward recall.

iv. Percentage of students "passing" a posttest, where the pass mark could be 100%, or could vary as in i.

v. Estimates of creativity, originality, success at problem-solving for higher-level learning (Gagné, 1970; Bloom, 1956).

Task Description. The developer must identify what discreet tasks the learner must be able to perform in order to achieve overall success. This breaking down of instruction into its sub-parts must be carried out for all types of instruction. It is wrong to think of the word "task" as applying just to motor skills. Cognitive content can be broken down into intellectual tasks. The task description provides both a framework around which the developer builds instructional materials and activities, and also a check-list that is used in field-testing.

For psychomotor skills, a task description can be carried out by thinking through the task, and writing down on a sheet of paper every step that has to be completed. It is best to err on the side of being too detailed at this point. It is easier to remove steps from the list than to add them later. Better still is to have a person who can do the task go through it, while writing down everything that the person does. This way it is less easy to overlook steps when you think it through. For cognitive skills, working through the task yourself and consulting others are about the best you can do to compile your list of sub-tasks. You cannot see what is going on inside someone elses head (though this can be deduced fairly accurately in some instances, such as solving mathematical problems, or translating into a foreign language, since each step or word produces something on paper). More complicated and empirical methods exist for the analysis of cognitive tasks; but they are best left alone by all but the most highly trained developers.

The result of the task description is a list of smaller tasks in sequence. These sub-tasks form the basis for the evaluation of the effectiveness of instruction.

Testing. To test the effectiveness of instruction in motor skills, the developer must actually observe the learner working through the series of tasks, and note on a score sheet whether each step is successfully completed or not. Learners should work through the instructional material and associated activities, and should then be tested individually on the skill. All that it is necessary to record is whether the learner was successful or not in completing the task. (A typical checklist for testing a motor skill -- laminating a picture -- is given in figure one).

Figure 1 About Here

If the learner becomes completely stuck, the examiner can help so that the learner can move onto the next step in the sequence. A note should be made of the examiner's intervention.

Testing cognitive skills will usually involve a written test to be completed after instruction. Remember, this is formative, not summative evaluation. It is therefore necessary to test not just the cumulative effects of instruction on the learner's knowledge and ability, but the successful acquisition of each sub-part of the total skill as it is encountered along the way. As a general rule, there should therefore be at least one question on each cognitive sub-task identified in the task description. For example, if the objective of the instructional unit is for the learner to be able to describe why an airplane flies, and the task description has identified the understanding of vector diagrams and of airflow over a curved surface as tasks leading to an understanding of the whole process, then the learner should be asked at least one question on vector diagrams and one on airflow as well as general questions on how

planes fly. The reason for this is that if you just had a learner describe how planes fly, and the learner was unable to do so, then you would not know if it was because there was a lack of understanding of vector diagrams or of airflow. You would know that instruction had not been successful, but you would not know where the learner had had difficulty.

Tests of cognitive abilities may be of any kind -- multiple choice, short answer, or even essay, provided that instructions are given to the learner which require an explanation of all pertinent sub-tasks. A word of warning, though. It is difficult to construct good multiple-choice tests, which are valid, reliable and discriminate well between those who have learned the material and those who have not.

Sample Size. You are field-testing, not conducting a sophisticated psychological study. The early rounds of field-testing should identify the most serious weaknesses and the most obvious mistakes. These should become apparent in the behavior and in the written answers of the first few of your learners. On some psycho-motor skills, it is not necessary to test more than three or four learners in the first round of field-testing, increasing the number to half a dozen or as many as eight in subsequent rounds. The same is true for cognitive skills. Written answers from four learners will give you a satisfactory first estimate of the success of your instructional materials. Keeping the numbers small will also leave sufficient learners in the pool of potential "subjects" for subsequent rounds of testing and evaluation.

Interpreting Scores. There is no need to submit learners' scores to elaborate statistical analysis. Looking at individual learners' scores is quite sufficient to identify where problems lie. Figure two shows a

score-sheet of errors committed by just two learners during field testing

Figure 2 About Here

of the laminating unit mentioned above. Each learner made two attempts at the task. Both missed step eight the first time around (cutting off excess laminating film), and one of them became quite confused over the whole sequence for preparing the film to cover the illustration during the second trial. This should be enough to indicate to the developer that the instruction was not clear about removing excess film and preparing the film to lay over the illustration. Testing another fifty learners would not add much more information about this.

The interpretation of scores on written answers of whatever kind works in the same way. Incorrect answers to questions based on sub-skills upon which an understanding of the whole is based point to where the instruction is not as effective as it might be. The inability of more than one learner out of four to answer correctly a question on vector diagrams would suggest to the developer of the unit on flight that instruction was weak at this point.

Revision. The developer now looks at the faulty sections of the instructional materials that have been identified in the testing. Revision can be referred back to any point in the development process. Perhaps the section of videotape showing excess laminating film being cut off was terribly out of focus, and just needs to be re-shot. Perhaps the producer did not think it was terribly important, and showed it far too quickly to be understood. Perhaps it is not really necessary to cut off excess film at this point. Or, perhaps the understanding of vector diagrams is

asking too much of the learners, and the objective is not attainable. Whatever revisions are needed, the developer should make them, and submit the materials and activities to a second round of testing. The test-revise cycle is continued until the pre-established criteria have been met.

Evaluating the Feasibility of Instruction

Criteria. Typical criteria for assessing the feasibility of implementing instructional materials and activities are:

i. Cost. Can the "client" afford to purchase and maintain the instructional system?

ii. Available equipment and facilities. Normally, cost and available facilities are known before development begins, so these constraints would be taken into account before field-testing.

iii. Personnel. Are the teachers who will be using the instructional materials and activities trained to use them?

iv. Cost-effectiveness. Maybe learning takes place and the client can afford the system. But it is still relatively expensive for the amount of learning that takes place.

v. Difficulties in carrying out instruction. Maybe the instruction is successful, but requires an inordinate amount of time, or preparation.

Evaluation. The evaluation of the instructional materials and activities against these criteria can be done informally. Cost and facility factors can be assessed quite simply by questioning teachers and administrators. Personnel skills and difficulties in carrying out the instruction can be estimated by direct observation of teachers and learners

trying the materials out in classrooms. Maybe teachers need inservice training to implement the instruction, and maybe the activities and materials should be organized in different ways so that they are easier to carry out and use. The observer in the classroom should take notes of any logistical or other problems that occur so that modifications can be made to overcome them.

The above stresses the importance of doing at least some of your field testing in classroom settings if it is at all possible. The testing of individual learners, described in the previous section, needs to be supplemented by try-outs under conditions as close as possible to those in which instruction will be implemented. It is only through this kind of observation that certain problems can be detected. A set of self-instructional units, produced by the Learning Technology Unit at the University of Calgary (Learning Technology Unit, 1979) ran into a problem which observers noted. Materials designed for use by one learner were in fact being worked through by several learners at a time. As a result of this a division of labor occurred and no one student performed all of the required tasks. The reason for this was a higher enrolment in a lab. component of a course than had been anticipated, and the problem was overcome by duplicating stations in the lab. area. The infeasibility of self-instruction under these circumstances might have gone undetected had not part of the field-testing involved direct observation in the classroom setting.

Evaluating the Affective Component

Criteria. Instruction should reach acceptable levels in the following areas:

i. The students should like the activities and materials. If twenty percent or more indicate dislike, revisions should be made.

ii. The materials should capture and hold the interest of the learners. If twenty percent or more indicate boredom, the materials should be made more interesting.

iii. The materials should provide encouragement (feedback, reinforcement) to the students. If they feel a lack of encouragement or of direction, revisions should be made.

Testing Affect. More is going to be said in this session about measuring attitudes. So let us look at a few simple ways of finding out learners' affective reactions to instruction. Urie Bronfenbrenner (1976) has pointed out that the simplest way to find out if learners like the instruction, (or experimental treatments) they receive is to ask them. Every developer should at least do this. As learners run through the instructional materials and are tested, it is not difficult to ask them a few questions, such as: "Did you find this interesting or boring?" "Was it too easy or too difficult?" "Would you like to be taught more lessons this way?" Responses to these informal questions can be noted for future reference. Interest can be increased in revisions by speeding up, or even just changing the pace of instruction; by switching from a verbal to a visual form of presentation, or vice versa; by inserting an example into an expository segment of instruction, and so on.

Further refinements to these simple techniques can be made. Information can be obtained by having each learner, after working through the materials, rate statements on Likert scales. Typical statements would be: "I found the material interesting," rated from "strongly agree" to

"strongly disagree". Information about the learner's likes and dislikes can, in this way, be quantified, and sometimes this is an advantage. Or an affective component can be built into an observer's checklist (Thiagarajan, 1971). Here, in addition to a check mark or a cross against each step, to indicate success or failure, other symbols can be used. An exclamation point might indicate boredom, an asterisk, enthusiasm. In this way, revisions can be directed to specific portions of the instructional materials and activities, and motivational devices can be built into the parts of the materials where they are most needed.

It is also important to test the affective reactions of the teachers, or of whoever it is who will be present when the materials are used. The students may learn what they are supposed to, and they may like the materials very much. But if the teachers do not approve of the way the learners are to be taught, or if they are not enthusiastic about what the developer has produced, then it is likely that success will be difficult to achieve. Again, teachers can simply be asked their opinion, or can rate statements on scales. The developer needs no more information than this.

Summary

A Bare Minimum. For someone who is not accustomed to doing instructional development, and who is faced with the task of developing and producing materials and activities for a particular purpose, the foregoing has provided some suggestions as to how the developer's product can be tried out. The purpose of that trial is to provide information that will permit sensible revisions to be made to materials and activities with a minimum of fuss. To achieve this, it is not necessary for the casual

developer to carry out an elaborate, empirical, statistical evaulation. But the developer should at least carry out the following:

i. Set criteria for acceptable learner performance.

ii. Determine criteria for feasibility of implementation.

iii. Determine what will be an acceptable affective rating.

iv. Perform a task description.

v. Prepare an observer's checklist, or written test, depending on the content, (motor skill or cognitive), to evaluate the effectiveness of the instruction for each step as well as for overall ability.

iv. After giving the test, or observing performance, ask the learners if they liked the instruction, if they found it hard or easy.

vii. Quiz teachers and administrators on the feasibility of implementing the instruction.

viii. Ask teachers if they liked the instruction, and if they think it is appropriate for the intended learners.

ix. Carry out revisions, after a try-out with four to six learners.

x. Try to field-test the instruction, or some part of it, in a real classroom setting.

Useful Elaborations. For those with a little more time and initiative, a few additional features can be added to field-testing which make the task easier, more empirical and more objective:

xi. If you have skill in the matter, write multiple choice tests so that the distractors perform a diagnostic function, i.e. b. is the right answer, but chosing c. indicates that the learner has a different kind of difficulty from the learners who chose a.

xii. Develop a coding system on observation checklists to indicate reactions, such as boredom and enthusiasm, beyond just right and wrong.

xiii. Use rating scales for affective responses.

With these few simple tools, any teacher or anyone who is not a trained instructional developer, should be able to carry out simple field-testing of instruction and revise it accordingly. While development of large and complex instructional systems requires far more elaborate field-testing, and while even field-testing of simpler systems and materials can be more complicated, educational technology is not inherently esoteric. Galbraith's definition of technology as the application of knowledge to practical tasks is not difficult to understand, nor to apply. And it is in the application, through field-testing, of theory to practice, that educational practitioners can, in their everyday activities, become educational technologists.

References

Bloom, B.S. *Taxonomy of Educational Objectives. Handbook I: Cognitive Domain.* New York: McKay, 1956.

Bloom, B.S., Hastings, J.T. and Madaus, G.F. *Handbook on Formative and Summative Evaluation of Student Learning.* New York: McGraw-Hill, 1971.

Bronfenbrenner, U. The experimental ecology of education. *Educational Researcher,* 1976, 5, 5-15.

Committee on Definition and Terminology. The field of educational technology: A statement of definition. *Audiovisual Instruction,* 1972, 17 (October), 36-43.

Gagné, R.M. *The Conditions of Learning.* New York: Holt, Rinehart and Winston, 1970.

Gagné, R.M., and Briggs, L.T. *Principles of Instructional Design.* New York: Holt, Rinehart and Winston, 1974.

Galbraith, J.K. *The New Industrial State.* Boston: Houghton-Mifflin, 1967.

Kaufman, R.A. *Educational System Planning.* Engelwood Cliffs: Prentice-Hall, 1972.

Learning Technology Unit. *Evaluation of Practicum Self-Instructional Units.* (L.T.U. Research Report No. 2). Calgary: Learning Technology Unit, 1979.

Thiagarajan, S. *The Programming Process: A Practical Guide.* Worthington Ohio: C.A. Jones, 1971.

FIGURE I

DRY MOUNT PRESS: Task Description /SEALAMIN

TASK	YES	NO
1. Set thermostat at 225° F.		
2. Turn on		
3. Set tacking iron to 'medium'		
4. Select appropriate illustration		
5. Select Sealamin large enough to cover		
6. Dull side down		
7. Smooth out Sealamin		
8. Cut off excess Sealamin		
9. Place materials in newsprint sandwich		
10. Place sandwich in press		
11. Materials in for 30 seconds		
12. Check for bubbles, wrinkles		
13. (Reinsert if wrinkled)		
14. (Final trim)		

FIGURE II

MATERIALS PRODUCTION LAB

Tasks	Trial #	Task Errors with Two Subjects 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Dry Mounting	1			1		1	1			1				2						
	2									1				1						
Chartex	1			1						1										
	2							1		1										
Spirit Duplication	1														1					
	2					1	1													
Laminating (Sealamin)	1								2											
	2						1	1	1	1		1								
Laminating (Contact paper)	1							2												
	2		1							1	1	1	1	1						

NOTE: For task descriptions see appendices